John von Neumann Institute for Computing

A Statistical Approach to Deriving and Analyzing a Propensity Scale for Predicting Exposed Transmembrane Beta Barrel Residues from Protein Sequence

S. Hayat, Y. Park, V. Helms

published in

From Computational Biophysics to Systems Biology (CBSB07),
Proceedings of the NIC Workshop 2007,
Ulrich H. E. Hansmann, Jan Meinke, Sandipan Mohanty,
Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **36**, ISBN 978-3-9810843-2-0, pp. 137-140, 2007.

http://www.fz-juelich.de/nic-series/volume36

A Statistical Approach to Deriving and Analyzing a Propensity Scale for Predicting Exposed Transmembrane Beta Barrel Residues from Protein Sequence

Sikander Hayat, Yungki Park, and Volkhard Helms

Center for Bioinformatics, Saarland University, 66041 Saarbrücken, Germany
E-mail: {s.hayat, volkhard.helms}@bioinformatik.uni-saarland.de

In the current study, we implement an algorithm to analytically derive a novel propensity scale for the Transmembrane Beta barrel (TMB) residues to be exposed to the lipid bilayer. Since it is very difficult to experimentally determine their 3D structures and given the fact that they perform several important functions in the cell proteome of both gram-negative bacteria and eukaryotes, it is imperative to develop *in silico* methods for the modeling of their 3D structure. The algorithm previously described by us for Transmembrane Alpha helical proteins takes into account the evolutionary conservation and frequency profile to derive a positional score for a given Transmembrane residue. The scale, based on ridge regression, is derived such that the positional score for a given residue is maximally correlated with its relative solvent-accessible surface area (rSASA) value. A leave-one-out test with known structures demonstrates the correlation coefficient between the observed and predicted rSASA values to be around 0.52. Analysis of scales derived for both the interface and the hydrophobic core of the TMBs provides interesting insights into structural aspects of TMB residues.

Keywords: ridge regression, TM Beta Barrel proteins, frequency profile, propensity scale

1 Introduction

TMBs, composed of antiparallel Transmembrane beta strands and connected by soluble loop regions are inserted into the outer membrane (OM) of gram-negative bacteria, mitochondria and chloroplasts, where they perform a variety of functions, including passive transport of ions and small hydrophilic molecules, membrane anchoring and a role in bacterial pathogenicity[1–4]. The lipid-facing surfaces of the barrels are composed of hydrophobic residues and the residues facing the interior of the barrel are mostly polar residues[5] to facilitate transfer of small solute molecules. Accurate structure prediction of TMBs still remains a challenge, as Transmembrane protein structure can not be easily studied using X-ray crystallographic and NMR methods. As of now, no propensity scale exists that can account for the propensities of TMB residues to be exposed to the lipid bilayer. In this study, we implement an algorithm previously established by us[7] for deriving a propensity scale for Transmembrane Alpha helical residues to formulate two propensity scales for the Beta barrel residues to be exposed to the lipid bilayer at the hydrophobic core and at the interface region, respectively. Briefly, the algorithm tries to maximize the correlation coefficient between the observed and the predicted rSASA values for a given residue. The rSASA value of a residue describes the extent to which that particular residue is exposed to the lipid bilayer. Previously, it has been observed that polar residues tend to be buried inside and hence are less exposed to the bilayer in the hydrophobic core region[5]. The same is not true for residues at the interfaces, where the distinction between the exposure patterns is influenced by the changing nature of the lipid environment. It is also known that

the more conserved a residue is, the more it is structurally and/or functionally important for the protein. Consequently, the frequency profile of the 20 naturally occurring amino acids and their positional conservation Index is used as the input vector.

2 Materials and Methods

A non redundant data set of known TMB structures was compiled primarily based on the list provided by Tusnady *et al.*[6]. The final dataset comprises of 25 protein chains with 2305 and 1195 TM residues in the hydrophobic core and interface region, respectively. The hydrophobic region was derived from the OPM database[10]. The classification of each residue as being buried or exposed to the lipid bilayer was based on its rSASA value. As previously described by us[7], the SASA values were calculated with the VOLBL program suite employing a probe radius of 2.2 Å. To prevent the residues lining the internal cavity of the protein from being classified as exposed, capping with dummy residues was performed for both the interfaces and the hydrophobic core. SASA values were normalized to generate rSASA values by considering the SASA values for each amino acid X in the context of the tri-peptide G-X-G. The positional frequency for a given residue was calculated using AL2CO program suite. The complexity parameter for ridge regression was found by employing 10 fold cross validation. Calculations performed using R yielded 0.691 and 0.301 as complexity parameters for hydrophobic and interface regions, respectively.

3 Results and Discussion

3.1 Implementation of an Optimum Propensity Scale for TMBs

	PHE	MET	TRP	ILE	VAL	LEU	ALA	PRO	ASP	GLU
HMOB	0.086	-0.086	0.030	0.023	0.077	0.246	0.056	0.068	-0.062	-0.055
IMOB	0.096	-0.040	0.094	0.044	0.019	0.100	-0.017	0.030	-0.057	-0.059
MO	-0.010	-0.230	-0.030	0.050	0.020	0.020	-0.090	-0.100	-0.270	-0.200
	CYS	ASN	GLN	THR	TYR	SER	GLY	HIS	ARG	LYS
HMOB	0.047	-0.061	-0.060	-0.035	-0.020	-0.075	-0.026	-0.026	-0.062	-0.063
IMOB	-0.050	-0.043	-0.031	-0.006	0.078	-0.063	-0.038	0.020	-0.038	-0.041
MO	-0.160	-0.230	-0.220	-0.180	-0.150	-0.190	-0.180	-0.240	-0.210	-0.100

Table 1. The propensity scales for transmembrane Beta-Barrel proteins. HMOB = Hydrophobic MO Beta, IMOB = Interface MO Beta, MO = MO scale for transmembrane alpha helical proteins. The standard deviation of individual propensity values with leave-one-out test was found to be less than 0.008 in all cases.

As can be seen in Table 1, the propensity values for TMB residues to be exposed to the bilayer are smaller in magnitude than their counterparts in Transmembrane Alpha helices. This could be due to the less hydrophobic exterior of TMBs, which is necessary for their translocation via the inner membrane[8]. The affinity for hydrophobic residues to be exposed to the hydrophobic environment of the lipid bilayer is in concert with the experimental results[4].

	0.00^a $(0.44)^b$	0.01 (0.37)	0.02 (0.35)	0.03 (0.34)	0.04 (0.33)	0.05 (0.33)
HMOB CC(0.51)	67.2	73.1	75.0	75.8	76.0	77.4
+CI CC(0.52)	69.9	76.3	78.3	78.7	79.1	79.7
	0.00^a $(0.49)^b$	0.01 (0.45)	0.02 (0.44)	0.03 (0.42)	0.04 (0.41)	0.05 (0.39)
IMOB CC(0.48)	76.6	76.6	76.0	76.6	77.0	77.1
+CI CC(0.49)	77.2	78.3	77.7	78.0	78.8	78.7

Table 2. Performance comparison of the derived scales at different rSASA values in the hydrophobic core region. Entries reflect the accuracy of prediction, which improves with the inclusion of conservation index as an input parameter. a = Threshold rSASA, b = fraction of exposed residues, CC =Absolute magnitude of Correlation coefficient between the observed rSASA and the computed positional score, ACC = Accuracy of correct prediction in percentage, CI = Conservation index.

3.2 Performance Comparison of Propensity Scales

A leave one out test was conducted to measure the performance of the propensity scales. The performance of the particular scale was assessed by analyzing the correlation between positional score and the observed rSASA values and by the corresponding prediction accuracy. A support vector machine in R was used to implement the test. As depicted in Table 2, the choice of the cutoff value for the rSASA value is a major factor when it comes to predicting the burial status. In the current study, this cutoff was objectively chosen based on the SVM. The table also shows that inclusion of conservation indices as an input parameter enhances the performance of the derived scales. The overall weak correlation between the observed rSASA and the computed positional score suggests that the lipid protein interaction might not be the only factor involved in protein insertion and folding mechanism[8].

3.3 Comparative Analysis and Correlation with Other Scales

	Scale (Reference)	HMOB	IMOB
Hydrophobicity	KD (Kyte *et al.*, 1982)	0.68	0.46
	EIS (Eisenberg *et al.*, 1984)	0.63	0.61
	GES (Engelman *et al.*, 1986)	0.54	0.51
	WW (Wimbley *et al.*, 1996)	0.56	-
	Hessa (Hessa *et al.*, 2005)	-0.58	-0.52
Size	Bulkiness (Zimmerman *et al.*, 1968)	0.53	0.74
Packing	Partial specific volume (Cohn *et al.*, 1943)	0.60	0.59
Others	KPROT (Pilpel *et al.*, 1999)	0.61	0.30

Table 3. Correlational analysis with other scales. The propensity scales derived here shows weak correlation with other scales.

As shown in Table 3, one method to discover the other factors involved in TMB folding could be to find other scales that strongly correlate with the scales derived here. As expected, the hydrophobicity scales show a weak correlation with the propensity scales derived here, which can be atributed to the less hydrophobic exterior of TMBs.

4 Conclusion

The current study successfully implements the MO algorithm for Transmembrane Beta barrel proteins. The propensity scales for both the hydrophobic core and the interface region are presented. The derived scales confirm the less hydrophobic exterior of the TMBs and are weakly correlated with hydrophobicity scales. A further analysis of the scale in terms of principal components, distance of the residue from the lipid bilayer core and more advanced statistical methods need to be employed to fully understand the insertion and the folding mechanism of the TMBs in an analytical way. Development of a reliable predictor for the burial status of TMBs can be used to impose additional constraints on starting template models while performing *ab initio* structure prediction[9]. Further analysis of TMB residue propensities might provide important insights into the evolution of the mitochondrial OM and the development of its protein biogenesis system[8].

References

1. R. Koebnik, K. P. Locher, and Van Gelder P., *Structure and function of bacterial outer membrane proteins: barrels in a nutshell*, Mol. Microbiol. **60**, 1-4, 2000.
2. G. E. Schulz, *Porins: general to specific, native to engineered passive pores*, Curr. Opin. Struct. Biol. Mol. **6**, 485-490, 1996.
3. A. Pautsch, and G. E. Schulz, *Structure of the outer membrane protein A transmembrane domain*, Nat. Struct. Biol. **5**, 1013-1017, 1998.
4. G. E. Schulz, *The structure of bacterial outer membrane proteins*, Biochim. Biophys. Acta. **1565**, 308-317, 2002.
5. L. K. Tamm, H. Hong, and B. Liang, *Folding and assemble of beta barrel membrane proteins*, Biochim. Biophys. Acta. **1666**, 250-263, 2004.
6. `http://pdbtm.enzim.hu/` G. E. Tusnady, Z. Dosztanyi, and I. Simon, *Transmembrane proteins in the Protein Data Bank: identification and classification*, Bioinf. **20**, 2964-2972, 2004.
7. Y. Park, and V. Helms, *On the Derivation of Propensity Scales for Predicting Exposed Transmembrane Residues of Helical Membrane Proteins*, Bioinf. **23**, 701-708, 2007.
8. N. Pfanner, N. Wiedemann, C. Meisinger, and T. Lithgow, *Assembling the mitochondrial outer membrane*, Nat. Struct. and Mol. Biol. **11**, 1044-1048, 2004.
9. L. Adamian, and J. Liang, *Prediction of transmembrane helix orientation in polytopic membrane proteins*, BMC Struct. Biol. **6**, 2006.
10. M. A. Lomize, A. L. Lomize, I. D. Pogozheva, and H. I. Mosberg, *OPM: Orientations of Proteins in Membranes database*, Bioinf. **22**, 623-625, 2006.